Karlsruhe Institute of Technology

BACHERLORARBEIT

# Ontology Repositories and the Role Linked Data has on them

von
Helena Hibtes

eingereicht am 28. September 2012 beim
Institut für Angewandte Informatik
und Formale Beschreibungsverfahren
des Karlsruher Instituts für Technologie

Referent: Prof. Dr. Studer
Betreuer: Dr. Elena Simperl
Betreuer: Benedikt Kämpgen

Heimatanschrift:
Quintinsstraße 9
55116 Mainz

Schriftliche Erklärung:

Ich versichere hiermit an Eides statt, dass ich die vorliegende Bachelorarbeitselbstständig und ohne unzulässige fremde Hilfe erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie wörtliche und sinngemäße Zitate kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

| | |
|---|---|
| _____ | _____ |
| Ort, Datum | Helena Hibtes |

**Inhaltsverzeichnis**

## List of Illustrations

**Abbreviations:**

| | |
|---|---|
| **API** | applicationprogramminginterface |
| **BIBO** | BibliographicOntology |
| **BTCD** | Billion Triple Challenge Data |
| **CC** | Creative Commons |
| **CCBYSA** | Creative CommonsAttribution-ShareAlike 2.0 |
| **CKAN** | ComprehensiveKnowledge Archive Network |
| **CSV** | Comma-separatedvalues |
| **DAML** | DARPA Agent Markup Language |
| **DARPA** | Defense Advanced Research Projects Agency |
| **DCAT** | Data CatalogVocabulary |
| **DCMI** | Dublin Core Metadata Initiative |
| **DL** | Description Language |
| **FOAF** | Friend of a friend |
| **HTTP** | Hypertext Transfer Protocol |
| **IBM** | International Business Machines |
| **IEEE** | Institute of Electrical and Electronics Engineers |
| **JSON** | JavaScript Object Notation |
| **LGPLv3** | GNU Lesser General Public License Version 3 |
| **LOD** | Linked Open Data |
| **LOV** | Linked Open vocabulary |
| **MIME** | Multipurpose Internet Mail Extensions |
| **NeOn** | NetworkedOntologies |
| **OBO** | Open Biological and Biomedical Ontologies |
| **OIL** | Ontology Interchange Language |
| **OKF** | Open KnowledgeFoundation |
| **OMV** | OntologyMetadataVocabulary |
| **OR** | Ontology Repository |
| **OWL** | Web Ontology Language |
| **PDF** | Portable Document Format |
| **RDBMS** | Relational Database Management System |
| **RDF** | Resource Description Framework |
| **REST** | Representational State Transfer |
| **RRF** | Rich Relase Format |
| **SeCo** | Semantic Computing Research Group |
| **SKOS** | Simple KnowledgeOrganization System |
| **SOAP** | Simple Object Access Protocol |
| **SPARQL** | *SPARQL Protocol and RDF Query Language* |
| **SQL** | Structured Query Language |
| **TS-ORS** | topic specific open rating system |
| **UML** | Unified Modeling Language |
| **URI** | Uniform Resource Identifier |
| **URL** | Uniform Resourcelocator |
| **VOAF** | Vocabularyof a Friend |
| **VoID** | Vocabulary of Interlinked Data sets |
| **WWW** | World Wide Web |
| **XML** | Extensible Markup Language |

# 1 Introduction

The Semantic Web [BeHL01] aims to extend the World Wide Web (WWW) by making information machine-readable, so that beside humans, software agents can understand and interpret information, as well. In contrast to the WWW, which describes Web pages, the Semantic Web intends to describe resources. At this, a resource can be abstract or physical, for instance a book, a person, a place or a Web page, and is identified by a Uniform Resource Identifier (URI). URIs are strings of characters and name resources [CaSh06]. Additionally, further technologies are evolved. For example, representation languages such as the Resource Description Framework (RDF) and its serialization formats [LiMe11] and the Web Ontology Language (OWL) to describe additional information about resources, which is called metadata [DeAn12]. A further important component of the Semantic Web is ontologies. Ontologies represent knowledge by describing objects and enable to reason with those. Since the development of ontologies rises enormously, ontologies become too large trying to cover whole domains and searching for them constitutes a problem, so platforms are developed in order to simplify the reuse of ontologies. Those platforms are called ontology repositories and data catalogs respectively and are provided with services and elements.

Currently there are several ontology repositories such as BioPortal and Cupboard, but also data catalogs like the Data Hub and Dcat aiming to facilitate the reuse of ontologies and data sets. Therefore, elements and services, which enable access to data sets and ontologies, managing them and hence process them for the reuse, are determined, firstly. Within the scope of ontology engineering, it is evident that the use of Linked Data [BiHB09] improves those services and elements considerably. Linked Data, which aims to interlink structured data and uses Semantic Web technologies such as RDF, OWL, and URIs, ensure that data is not only available for humans, but also for machines, enabling them to process data automatically. Hence, it is illustrated how Linked Data influences the components of repositories, so that those can be used more efficiently. Based on this, Linked Data requirements of ontology repositories and data catalogs are determined, which are examined within a multiple case study. Therefore, representative repositories are selected and examined according to

those Linked Data requirements and if those are fulfilled. In a next step, a case study of vocab.cc is carried out. Vocab.cc is an open source project, to fulfill the function of searching and using Linked Data vocabularies. The case study is carried out according to the Linked Data requirements on repositories. Based on this, recommendations are made for vocab.cc in order to improve its workflow, so that reusing data is facilitated. Finally, an outlook is given, describing further developments regarding ontology repositories and data catalogs, which aim to provide their data sets as Linked Data.

## 2 Relatedwork

The efficient reuse of ontologies and therefore sharing information has a major role within the evolution of the Semantic Web. Therefore, researchers have developed different platforms, on which ontologies and data sets are stored in order to provide access to information. Several researchers have evaluated these platforms by describing the functionalities, which support the efficient reuse.

In 2001, Ying Ding and Dieter Fensel determine in their work Ontology Library Systems: The key to successful ontology reuse [DiFe01] elements, an ontology library must consist of managing, adapting and standardizing ontologies. Managing ontologies comprises of how to store ontologies according their accessibility, classification and the modularization. It requires that ontologies must be identified unambiguously and maintain die evolution of ontologies in terms of different versions in order to provide the latest version of an ontology. Adaptation of ontologies means that searching and editing need to be facilitated by providing different search capabilities, such as keyword-based and advanced search, and enabling editing and submitting ontologies. Additionally, deriving further information in order to infer consequences from ontologies, evaluating and verificating [TaAS10] those belong to adaptation. The third element standardization declares that ontologies should be represented with standardized languages, for example RDF and DAML+OIL. According to these elements, a survey is carried out in which several ontology libraries such as Webonto[1], the DAML ontology library system[2], the Ontology Server[3] and five other ontology libraries are evaluated. Based on this survey, the researchers concluded requirements for ontology libraries to improve managing, adapting and standardizing.

A further survey [LHLS04] is carried out according to Semantic Web technologies and their use in Web portals providing information. This survey goes a step further than the former survey by selecting Web portals depending

[1] http://projects.kmi.open.ac.uk/webonto/
[2] http://www.daml.org/ontologies/
[3] http://www.starlab.vub.ac.be/research/dogma/OntologyServer.htm

on three characteristics, namely that Web portals collect information and offer it to a community, so that users are able to share and exchange this information. Additionally, those functionalities are provided by Semantic Web technologies. An evaluation criteria catalog was developed in order to compare existing platforms, including functions such as editing, browsing, searching and providing actual versions of ontologies. The evaluation scheme is based on three layers, namely: the information access, which evaluates the user interaction with the individual portals; the information processing, which evaluates five process steps and the grounding technologies consisting of Semantic Web technologies, for instance ontologies and their management, reasoning and Semantic Web Services, and system technologies such as data management. The evaluation comprises of ten Semantic Web portals such as OntoWeb and the KAON portal. The evaluation leads to conclusions about limitations concerning the exploitation of Semantic Web technologies, for instance the disability to interoperate with related portals and not providing alignment technologies, the insufficient use of methods to provide the latest ontologies.

The following survey extends the related works by adding the Linked Data aspect. The elements and services provided by ontology repositories are expanded by including Linked Data abilities to improve those. It is assumed, that Linked Data facilitates retrieving ontologies and data sets by providing a multitude of tools and hence improves their reuse.Therefore, Linked Data requirements are determined which ontology repositories and data catalogs have to meet to be sufficient. Additionally, a new vocabulary catalog is presented in order to demonstrate how ontology repositories and data catalogs can work efficiently by using Linked Data if it meets the Linked Data requirements.

# 3 Background

## 3.1 Ontology Repositories

Since ontologies are a key component that represents knowledge, the number of ontologies rises. Finding appropriate ontologies and concepts is difficult, so that a system which simplifies the information retrieval by carrying knowledge is necessary. Such a system is comparable with a library, since libraries are carriers of knowledge. Therefore, the librarian system is applied to data as information becomes increasingly available in the WWW [LiMe11]. This system is called a repository and as a location for storage, comparable with archives, it contains collections of data.
To point out the characteristics of repositories in general, a proposal by R. Heery and S. Anderson [HeAn05] is given:

- content is deposited in a repository, whether by the content creator, owner or third party
- the repository architecture manages content as well as metadata
- the repository offers a minimum set of basic services e.g. put, get, search, access control
- the repository must be sustainable and trusted, well-supported and well-managed.

Those characteristics have been developed further in order to adapt them to ontology repositories. This means that within the context of the Semantic Web, an ontology repository is available for both humans and machines. As a container of knowledge, the ontology repository represents a knowledge base, which is enriched with additional information about the ontologies. Therefore machines are enabled to reason about the content. Further services, for example searching, rating and mapping ontologies, are later discussed.

### 3.1.1 Ontologies

The term ontology originates from Greek philosophy and is about the theory of being [DeAn12]. In information science, an ontology describes concepts within a domain and the relationships between those concepts [LiMe11]. Therefore the data model is able to sort and interpret information, so that machines can solve queries. According to Gruber [Grub93], "*an ontology is a formal, explicit specification of shared conceptualization*". This means that the concepts and relations of a shared and abstract model become machine-readable. Therefore, an ontology describes a knowledge domain by using terms, relations and inference rules. Ontologies are expressed using the formalized specifications, for instance the Resource Description Language[4][5](RDF) and the Web Ontology Language(OWL)[6][7][DeAn12].

### 3.1.2 Need of ontology repositories

Since the number of ontologies rises constantly [HaPG09], there is a need to organize them. Therefore, ontologies have to be stored at one place in order to organize them and support their reuse [DiFe01].
First of all, storing ontologies in an ontology repository enables sharing ontologies and concepts by its services and elements since those are available at one place. Additionally, there is a need for facilitated search and browsing methods in order to find appropriate ontologies and concepts. Often the proper ontologies are found much too late and it becomes difficult as search takes an enormous amount of time [TSVH10].
Constantly developing new ontologies does not represent the original intention of knowledge managing. Ontologies, as carriers of specified knowledge with an accurate description of information and its semantics and are envisaged to be reused [DiFe01].
Besides, the designed ontologies are often too extensive [Link11] because the knowledge is too wide to be captured by a single ontology. Consequently, the

[4] http://www.w3.org/TR/rdf-concepts/
[5] [Worl04]
[6] [Mcva04]
[7] http://www.w3.org/TR/owl-features

ontologies cannot be reused efficiently, as only a part of the ontology is commonly needed or rather desired to be reused. But having small and simple ontologies describing parts of a knowledge domain does in fact simplify their reuse. But for this purpose, finding them cannot be a problem again. On the other spectrum, exceedingly small ontologies are most difficult to find. A successful repository covers a wide field of knowledge [Link11]. Therefore, a repository should include a large number of small ontologies, which can be connected by interlinking them through mappings, so that a wider field can be described [TSVH10].

### 3.2 Data Catalogs

A data catalog is a collection of data sets. They are collections of data and contain information. Hence, one data set contains information about a certain topic, for example medical data.
Data catalogs are used in order to provide data concerning certain communities. The contained data sets are expressed in formal specifications, like RDF and the Extensible Markup Language (XML)[8]. In the context of open data, communities, especially government initiatives provide data, which is freely available, so that licenses or copyrights are not needed. The Open Knowledge Foundation (OKF)[9], which researches and promotes open knowledge, defines open in relation to data as:

*“a piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike”*.[10]

Such a data catalog is for example, Data.gov[11], the data catalog of the United States Government. The intention behind Data.gov is to provide machine-readable data sets easily for the public. Data.gov provides descriptions of metadata which is information about how to access the data sets. Through this

[8] http://www.w3.org/XML/
[9] http://okfn.org/
[10] http://opendefinition.org/
[11] http://www.data.gov/

data catalog, transparency should be reached by enabling the public to participate in government applications such as research and carrying out analyses. Additionally, users can suggest data sets that should be added.

## 3.3 Linked Data

### 3.3.1 The Linked Data Technology

Linked Data is about to make typed links between data from diverse domains. In doing this, it is neither important if the data is from different organizations nor if the data is handled within one organization. Technically seen, data is published in machine-readable formats and is explicitly defined. That means that humans and machines are able to explore the Web of data. By the typed links made between internal and external data, it is possible to find other related data. The links between arbitrary things are described in RDF [Bern11], [BiHB09]. Things from the real world get Uniform Resource Identifiers (URIs)[12] [SaCy08] labels. In order for all data published on the Web to belong to the global data space, there are four rules known as the Linked Data principles designed by Tim Berners-Lee [Bern11]:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things.

Using the Semantic Web means working with URIs [Bern11]. Therefore, it is important to give real-world things a URI in order to identify them. It is better to provide these things with a HTTP URI because the HTTP name lookup is a very powerful and complex set of standards. In order to gain the desired information, URIs must be used efficiently and in order to be able to do that, it is imperative that RDF or SPARQL is used for query service. This is helpful because the

[12] http://www.w3.org/TR/uri-clarification/

information is available on the Web as Linked Data and not archived. Making links is necessary to connect data and to have a web of connected things. Links to other URIs provide finding different things, similar to the hypertext web.
As the Linked Data principles already illustrate, URIs and HTTP are the two most significant technologies. When things can be identified by URIs in conjunction with the HTTP protocol, they can be found easily. That is because HTTP protocol provides a simple mechanism for retrieving resources or descriptions. This facilitates the publication of data and the addition of it to the global data space [HeBi11].

### 3.3.2 Publishing Linked Data

Publishing Linked Data involves:

1. Choosing URIs: there is the opportunity to choose between two patterns: 303 URIs and hash URIs [BiHB09]. The latter identifies real-world things, which are separated from the remaining part of the URI and cannot be retrieved directly. The part separated by the hash symbol # is called fragment identifier [HeBi11]. The 303 URIs identify Web documents directly. The server responds to the client with the HTTP status code 303 See other. After this, the client gets the Web document, which describes the real-world object [HeBi11], [SaCy08].

2. Providing RDF links: they enable browsing the Web and therefore, finding further resources. By following the links the user gets more information about a certain topic [BiHB09].

3. Adding metadata: because it reports about the quality of data and extends the usability for the user. This is an important factor when it is about deciding whether the resource and its data is trustful or not. Useful clues are information about the creator, the method used, and when it was created [BiHB09].

### 3.3.3 Linked Open Data

The largest application is the Linked Open Data (LOD) project. The idea is to offer all data stock which is used by many users of the WWW. “Open” stands for free usage and distribution as before defined by the OKF. The goal is to make common used data from sources about geography, science and books, just to name a few, available without any restrictions in order to extend the Web. Consequently, added value is gained. There is also open government data for instance, the official Web site of the United States Government[13], which is integrated in the project. At first, existing data sets are identified. While doing this, the offered data is published under an open license. A license, which is permission to use the data, should be valid worldwide to keep in line with the community spirit of the LOD movement. These data sets are changed in RDF and after this, published on the Web. Data sets from different sources are connected by RDF links. Therefore, the user can navigate through different, but related data. Many important organizations like IBM[14] or the IEEE[15] joint the Linked Open Data project. This illustrates its immense growth since its establishment in Janurary 2007 [BiHB09], [HeBi11].

The first diagram illustrates the evolution of the Linked Open Data project between May 2007 and July 2009. The second diagram illustrates the state-of-the-art of the LOD cloud.

[13] http://www.data.gov/
[14] http://www.ibm.com/us/en/
[15] http://www.ieee.org/index.html

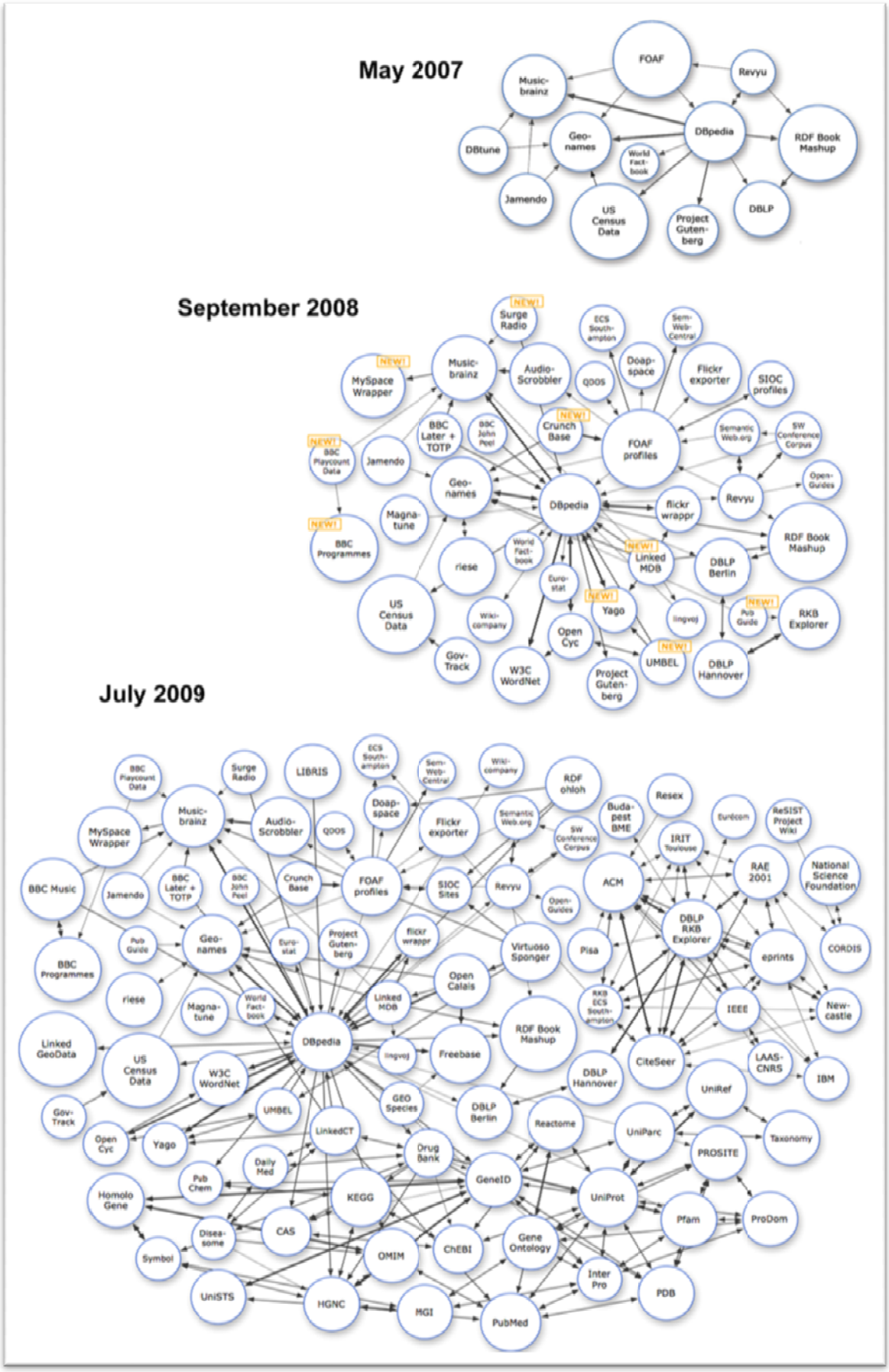



**Illustration 1[CyJe11]**

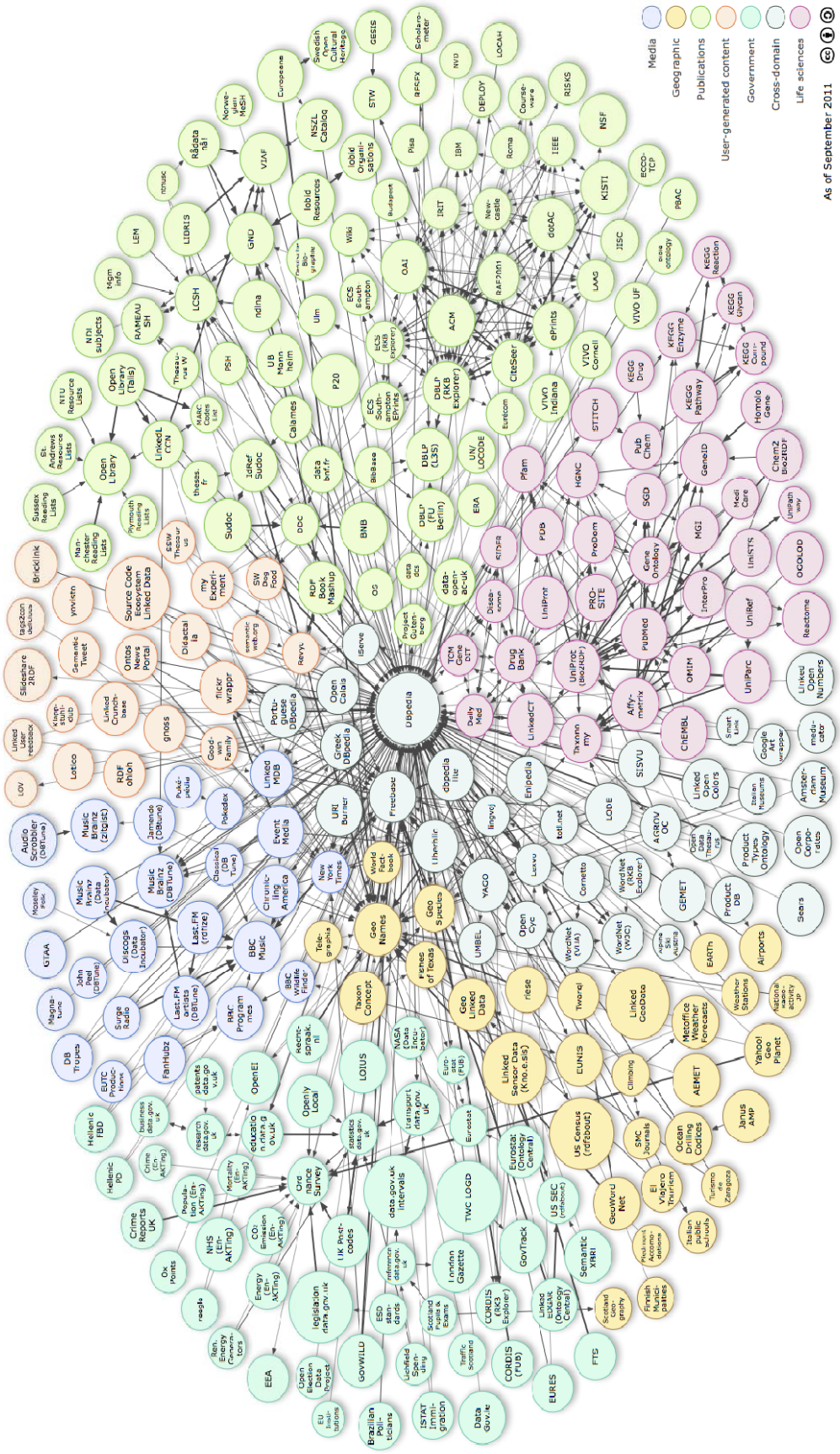



**Illustration 2[CyJe11]**

# 4 Study design

The following chapter describes the structure of the multiple case study, which will be carried out on representative ontology repositories and data catalogs. After that, a case study applied on vocab.cc[16] follows. Vocab.cc is an open source project. It was designed is for searching and using of Linked Data vocabularies.

## 4.1 Multiple case study

In the multiple case study the two following problems are centered:

1. Which requirements have Linked Data on ontology repositories?
2. In what way are the requirements satisfied by the selected ontology repositories?

In order to discuss the second problem, the first one has to be addressed. Therefore, requirements concerning Linked Data and the way to publish it correctly are listed and further explained. On the basis of this, the representative ontology repositories and data catalogues are examined more closely. This means that every single ontology repository and data catalogue respectively, are going be examined towards the requirements, which are determined in the first part. The results are listed in a table, with which a clear overview is provided.

After this, a discussion follows in which the results are used in order to take stock of the present situation regarding ontology repositories and data catalogues. It will be discussed how far the selected objects correspond to the requirements, which were determined in the first part of the multiple case study.

A brief description of the representative ontology repositories and data catalogues follows.

### 4.1.1 BioPortal

BioPortal[17] is an ontology repository providing biomedical ontologies and associated data resources. Created by the National Center for Biomedical Ontology, BioPortal is an open repository. It is based on biomedicine in order to express biomedical data and enable computers to read it. The user can access BioPortal via Web browsers and Web services. BioPortal supports browsing,

[16] http://www.vocab.cc/
[17] http://bioportal.bioontology.org/

publishing, searching and visualizing ontologies as well as mapping, commenting on them and reviewing. The latter supports the community character by enabling the user to participate {Salvadores 2009 #44} {Noy 2009 #18} {Noy #46} {LePendu 2010 #45}.

### 4.1.2 Cupboard

Cupboard is part of the NeOn[18] project, which offers ontologies a complete infrastructure to a community, where the ontology repository enables publishing, sharing and the reuse of ontologies. Cupboard provides access via the Web browsers Safari and Firefox. Cupboard develops technologies aiming to give better answers to queries that are integrated on a platform. Therefore users as well as publishers are supported. Besides this, the stored ontologies are available for applications. Cupboard offers every user that develops their own space. It contains the uploaded ontologies and the related ones, added metadata and reviews. With this in mind Cupboard represents not an exemplary repository, but a platform which stores several repositories. In addition, Cupboard provides features that facilitate sharing and reusing these ontology spaces [AqLe09], [AqEL09], [LeAE10].

### 4.1.3 Watson

Watson[19] is a search engine for Semantic Web documents. It offers three services, which facilitate the access to semantic data. Since a lot of semantic documents are distributed over the Web, Watson represents a gateway for the Semantic Web and is not only an adaptation of regular Web techniques but is construed for semantic applications. Therefore Watson considers explicit relations between semantic data and the implicit ones. Besides this, semantic quality plays a big role. Therefore data needs to be validated, indexed and ranked [ABGS11], [ASDB07].

[18] http://www.neon-project.org/nw/Welcome_to_the_NeOn_Project
[19] http://watson.kmi.open.ac.uk/WatsonWUI/

### 4.1.4 The Data Hub

The Data Hub[20] is an open data catalog and provides data sets from the Internet. The data catalog is community-based, so that users are able to add data sets and to search for them. Storing data sets and their visualization are enabled. The Data Hub uses the Comprehensive Knowledge Archive Network (CKAN)[21], which is an open-source data cataloguing software. CKAN is a data management system provided by the Open Knowledge Foundation (OKF), which stores data sets and provides metadata about those data sets.

### 4.1.5 OntoSelect

OntoSelect is an ontology repository providing ontologies on various topics. The platform enables different search options. In addition to searching an ontology, browsing is supported. The ontologies are analyzed in order to extract metadata and hence organized by indexing them with the referring information. OntoSelects reads the presentation languages RDFS, DAML and OWL. Besides this, developers can publish their ontologies and a ranking mechanism, depending on three criteria: coverage, structure and connectedness, which support the ontology retrieval [Buit04], [BuEi], [BuEi07], [BuED04].

### 4.1.6 Linked Open vocabulary (LOV)

The Linked Open vocabulary[22] provides vocabularies enriched with metadata and interlinked. Furthermore LOV enables to improve the visibility, usability and understanding of vocabularies. Designed and developed for users and editors, LOV represents a useful platform for developing, reusing and extending data sets. The LOV data set contains vocabularies formalized in RDFS and OWL ontologies. Another feature of LOV is the Vocabulary of a friend VOAF[23], which is a specification providing elements allowing the description of vocabularies used in the Linked Data Cloud. Besides that, LOV provides also usual metadata using for example, Dublin Core.

[20] http://thedatahub.org/
[21] http://ckan.org/
[22] http://labs.mondeca.com/dataset/lov/
[23] http://lov.okfn.org/vocab/voaf/v2.0/index.html

### 4.1.7 Data Catalog Vocabulary (DCAT)

Dcat is developed in order to describe data in data catalogs, so that dcat is applied, for instance, on the Linked Data catalog of the Australian government data[24] catalog [CyMP10] in order to improve the interoperability between different data catalogs. Is a standard RDF schema vocabulary, which reuses other vocabularies like the Dublin Core and FOAF [CyMP10].

### 4.1.8 ONTOSEARCH 2

ONTOSEARCH 2[25] is a search and query engine for ontologies. It reduces ontologies to the OWL DL-Lite representation language. ONTOSEARCH2 queries the Web to retrieve ontologies and stores a copy of them in a relational database [PaTS].

## 4.2 Case study on vocab.cc

In the case study applied on vocab.cc[26] the focus is laid on the problem:

1. How are the requirements of the multiple case study satisfied by vocab.cc?
2. Which recommendations for vocab.cc can be derive from the results, which were received from the multiple case study?

At this the requirements, determined for the multiple case study, are examined in order to give recommendations on vocab.cc.

Vocab.cc is an open source project and enables to lookup and search RDF vocabularies.

[24] http://data.gov.au
[25] http://www.ontosearch.eu/
[26] http://www.vocab.cc/

# 5 Motivation

In the following section the elements and services of ontology repositories and data catalogs are described. Afterwards a description about how Linked Data can improve those services and elements efficiently follows.

## 5.1 Elements and services of ontology repositories

The following requirements regarding the elements and services ontology repositories should provide, originate from different researchers and developers of ontology repositories. Based on the overall conditions by Ying Ding and Dieter Fensel [DiFe01], who determine first elements and services, which should be provided by ontology libraries, those are extended by the developers of BioPortal. Further initiatives are the Open Ontology Repository Initiative [BaSc09] and the Semantic Computing Research Group (SeCo)[27]. According to J. Hartmann, R. Palma and A. Gómez-Pérez [HaPG09], who define a generic ontology repository framework, which consist of a ontology repository and a management system,

"*An ontology repository (OR) is a structured collection of ontologies (schema and instances), modules and additional meta knowledge by using an Ontology Metadata Vocabulary. References and relations between ontologies and their modules build the semantic model of an ontology repository. Access to resources is realized through semantically-enabled interfaces applicable for humans and machines. Therefore a repository provides a formal query language*".

Based on this definition and the set requirements, a number of evaluation criteria for the multiple case study have been identified.

### 5.1.1 Information Access

The successful information access depends on the user interface, which enables the user to find appropriate ontologies and concepts by browsing the

[27] http://www.seco.tkk.fi/

repository through navigation bars. Here the visualization of the ontologies of importance. Besides this, a help-system providing useful advice about how to use the repository is provided facilitates.

- Search capabilities: Those should include full-text search, which can be refined by using mechanisms for filtering for example type, classes or properties, so that a multi-facet search interface is offered. For this, metadata is used. Additionally, an overview of the ontology or the data set is helpful in order to browse it and find out if it is relevant for the user's task [NoRM], [TSVH10].

- Visualization: is directly offered on the repository by different visualization tools. This includes for instance the presentation of data as graphs or of concepts as a hierarchy tree. Ontologies or data sets containing geographical information are presented using maps. Different formats, which are both machine-readable and suitable for humans, are provided, so that data is presented in a meaningful way [NoRM].

- Personalization: enables the individual visualization requests of users and improves therefore the information accessibility [HaPG09].

5.1.2 Knowledge Processes and Sevices:

The knowledge processes and services begin with the publication of ontologies and data sets on repositories. After submitting an ontology or data set, several steps follow in order to handle the knowledge, especially assessment mechanisms [HaPG09].

- Rating: measures an ontology's quality, coverage and usability inter alia. Since the user gives ratings, this is in fact a subjective task, but helps to guide other users. Reviewing is also an assessment mechanism, which is also a subjective since it is a report of one's experience with the ontology or the data set. Open Rating Systems constitute a solution to keep repositories efficient and effective, since any user can publish

ratings on the contained data. The meta-rating approach is provided because the quality of the ratings can vary. Therefore, the ratings have to be controlled as well so that the raters are rated [HaPG09].

- Evaluation: is in comparison to reviews objective, since the quality and adequacy of data is evaluated under consideration of specific goals. Therefore, metrics allow assessing simply data [HaPG09].

- Mapping: enable the interoperability between ontologies since those, or data sets, contain overlapping content, but have different structures, syntax and semantics. Mappings and alignments are depictions among ontologies and concepts, so that their elements are correlated. Hence the exchange of data and its integration is enabled [DeAn12], [NoRM].

- Reasoning: Inference rules support to derive knowledge form the data contained in a repository and therefore serve to process knowledge. Inference is the process of identifying automatically new relationships among data. This is done by deriving new data from the old one using queries. The gained information can be defined through vocabularies or sets of rules. Rules define mechanisms, which are able to retrieve and create new relationships based on old ones and hence improving the quality of information integration. Furthermore, inference mechanisms analyze the data to draw conclusions and manage it.

- Security: Since ontologies and data sets belong to intellectual property, licenses and copyrights are required. Also, clear access control and right management are required in order to enable access to the repository [HaPG09].

5.1.3 Organization:

- Lifecycle: A repository with lifecycle components documents the different stages of the evolution of an ontology, since those are dynamic [KDFO]. Hence mechanisms in order to update data are required [HaPG09].

Since ontologies are specifications of shared conceptualization [Grub93], there can be changes in the domain, conceptualization and specification [KDFO].

- Metadata: is essential since processing knowledge is based on it. Metadata provides data about data for example additional information about the domain or the author, so that information is further described and therefore, enables the communication with a machine or among machines. The reuse of data is provided because the retrieval and identification is supported and facilitated through metadata. Besides this, metadata can be used in order to maintain ontologies or data sets [DeAn12]. For the successful reuse of data, the creation, maintenance and distribution of metadata is supported by the Ontology Metadata Vocabulary (OMV)[28], for instance. OMV is a metadata standard, which offers terms and definitions to describe metadata [HaPG09].

### 5.1.4 Storage

- Indexing: Classifying ontologies facilitates searching, browsing and reusing them. In order to index data sets and ontologies there are different mechanisms. Data sets and ontologies can be indexed according their structure, subjects, and the different features offer for instance, according axioms and applications. Further classification categories are the relevance and components of ontologies and data sets respectively, for example domain affiliation and abstraction [KDFO].

- Identification: It is necessary to provide ontologies and data sets with unique names, identifiers and Unique Resource Identifiers, which facilitate searching since unambiguity is supported [KDFO].

- Language: A repository contains data, which is provided with different formal languages, for instance RDFS or OWL. The formal languages adjudicate on the expressivity of an ontology or data set**.** Those formal

[28] http://omv2.sourceforge.net/index.html

languageshave to be standardized in order to prevent misunderstandings. Here, controlled vocabularies help avoid homonyms and synonyms [KDFO].

## 5.2 Linked Data influence

Linked Data can be inserted in order to improve and facilitate the elements and services provided by ontology repositories and data catalogs by aiming to enable users to share structured data easily thereby supporting the reuse of data sets and ontologies. In the following how Linked Data can influence the knowledge workflows of ontology repositories and data catalogs is described.

### 5.2.1 Information Access

- Search capabilities: Using dereferenced HTTP URIs and interlinking data sets to access information is the first step. Ontology repositories and data catalogs should provide a set of crawlers, which process different formal specifications in order to find data sets by following the set links, which interlink data sets. There are also Linked Data search engines, for example Swoogle[29], which can be implemented into the ontology repositories and data catalogs since those come with their own set of crawlers on the Web [HeBi11].

- Querying: Using Linked Data enables sophisticated queries. Data sets and ontologies can be accessed via query endpoints. With a SPARQL Protocol and RDF Query Language (SPARQL)[30] endpoint can be applied in order to query data sets specified in RDF, RDFs and OWL Description Language (DL). SPARQL is also a W3C recommendation and enables queries of triple patterns [LiMe11]. This can be realized using for example the Jena API[31]. In order to query OWL, data sets operate using the OWL AP since it provides a query endpoint[32].

[29] http://swoogle.umbc.edu/
[30] [PrSe08]
[31] http://jena.apache.org/
[32] http://owlapi.sourceforge.net/

- Visualizing: The benefits gained from Linked Data in order to visualize data sets are the localization of errors is facilitated and users understand mechanisms without further knowledge since visual support is given for example, a graph visualization in the form of a scatter plots, which enable the comparison of two attributes, and histograms. Further visualization methods are mapped as well as and landscape views. There are two kinds of Linked Data browsers, which can be implemented, so that they enable different facets of visualization and therefore help the user to consume data. Those are Linked Data browsers, which provide text-based presentation and some, provide visualization options [DaRo11]. For instance, the Tabulator browser [BCCC06] belongs to the latter type. The Tabulator browser support less experienced user, since the visualization provides a wider range of presenting knowledge and is able to analyze large-scale data.

### 5.2.2 Knowledge processes and services

- Quality assessment: There are different types of quality assessment heuristics. The use of a certain heuristics is based on the desired quality factor. The content-based heuristics analyze the content of an ontology or compare it with a related one. There are outlier detection methods and spam detection methods. Another method assesses the quality by using metadata as a quality indicator. Based on the ratings users give based on different criteria of a data set, rating-based heuristics assesses the quality. Those quality assessing heuristics can be used in order to improve the search capabilities since data can be ranked according to the quality of properties and concepts or based on how often users select a certain ontology and concepts and properties respectively. The Linked Data search engine Sig.ma ranks data according the number of views of a data set [HeBi11].

- Mapping: Since different ontologies and data sets describe the same content, vocabulary mapping using Linked Data applications represents a useful solution in order to offer an integrated view. This occurs by

translation using vocabulary links like `owl:equivalentClass` mappings and statements like `rdfs:subClassof`. Those statements are published by the authors of the data and the maintainers, respectively. Consequently, the fourth Linked Data principle is applied, since links can be set between data from different resources [BiCH07]. Since RDFS and OWL do not support the fusion of two resources or splitting string values, expressive mapping languages, for instance the Alignment API, provide this [HeBi11].

- Reasoning: Linked Data requires providing data specified in RDF. Therefore data becomes machine-readable, which enables to generate additional information since machines are enabled to reason after querying [BiCH07]. Sending several queries, new relationships are retrieved. This enables to infer, so that new knowledge is produced or rather artificial intelligence is created and hence represents a key principle of the Semantic Web.
  Inference is the process of identifying automatically new relationships among data. This is done by deriving new data from the old one using queries. The gained information can be defined through vocabularies, which will be explained later, or sets of rules. Rules define mechanisms, which are able to retrieve and create new relationships based on old ones and hence improving the quality of information integration in order to expand the Web of data. Furthermore inference mechanisms analyze the data to draw conclusions and manage it after it.

### 5.2.3 Organization

- Lifecycle: Through the use of Linked Data, the lifecycle of an ontology can automatically be uploaded, so that always the newest version is stored. This is because of using URIs instead of other identifiers, which do localize the entity directly.

- Metadata**:** Linked Data enriches metadata by making statements about the provenance by being represented as RDF triples.

### 5.2.4 Storage

- Identification: Linked Data improves the identification of data sets by using HTTP URIs. Those are unique identifiers which describe directly the identified object or concept in contrast to URLs, which identify the address of the data set. Using the HTTP mechanism, those data sets are dereferenced. This means the data sets can be looked up directly [HeBi11].

- Language**:** Using the standardized format RDF and the serialization formats like RDF/XML[33], and RDFa[34] inter alia, structured data is available in the Web.

  .

[33] http://www.w3.org/TR/REC-rdf-syntax/
[34] http://www.w3.org/TR/xhtml-rdfa-primer/#using-rdfa

# 6 Linked Data requirements on Ontology Repositories

Based on the previous section, a number of requirements are identified, which ontology repositories and data catalogs should fulfill in order to work efficiently and therefore support the simple reuse of data. Since the previous section revealed, that Linked Data can be used in order to improve and facilitate the processes within a repository, the following requirements are based on the four Linked Data principles by Tim Berners-Lee [Bern11].

1. In order to access information, Web documents, concepts and properties should be provided with URIs in order to avoid duplications. Therefore, the entities of a data set can be identified by being already described by the URI. Besides, also real-world things should be provided with a URI [SaCy08], [HeBi11].

2. Provided with HTTP URIs, data sets can be looked up directly. The HTTP protocol mechanism provides the transmission of data within the WWW and therefore the communication with it [SaCy08]. These HTTP URIs have to be dereferenced in order to enable, beside humans, also machines to read data. A HTTP mechanism, which provides representations for humans and machines, is content negotiation [HeBi11]. It enables coordinating of the requested information in order to give the prevailing best representation for the individual client. Therefore for the server must distinguish whether the client is human or a machine. HTTP responses include a list of entities, which the client might get. For humans, the list contains entities, which are HTML because this representation is mostly preferred by humans. In contrast to this, the list for machines contains entities, represented in RDF. Hence the server can generate the representation [DeAn12].

3. The data should be represented by a standardized language in order to be machine-readable, which enables also machines, such as computers, to understand data. This includes the standard formats: RDF and its serialization formats RDFa, Notation 3, RDF/XML and Turtle [LiMe11].

This is also important when it comes to reasoning since more information or rather metadata according to data can be generated when also machines are able to reason [BiHB09].

4. Ontology repository and data catalogs should provide search, browse and index functions in order to retrieve data [HaPG09].

5. Querying endpoints supported by Linked Data can carry out sophisticated queries, which are structured data themselves. SPARQL is used as query language for RDF and its serialization formats. To handle the facilities, which RDF provides, SPARQL queries insist on triple patterns, similar to the RDF triples[35].
6. Since visualizing data is an important factor in order to browse and understand the data, Linked Data browsers and visualizing tools are needed [DaRo11].

7. Quality Assessment should be provided in order to support the user to select appropriate data, since evaluation represents reference about how a data set and an ontology respectively performs [HeBi11].

8. Ranking data while searching should not only result by matching the most keywords, but on the assessed quality for instance, ranking occurs according the most used concepts or the highest rated concepts [HeBi11].

9. Mappings between data should be used, since equal or similar data can be provided with different URIs. Mappings facilitate the retrieving process.

10. Metadata about the provenance of the data set should be provided since different HTTP URIs may identify the same entity [HeBi11].

[35] [PrSe08]

11. Metadata according to the security, by means licenses, needs to be provided, so that data is available legally [HeBi11].

12. Metadata to describe data and hence provide additional information, using vocabularies, should be used. The access to data is facilitated, consequently, since a common set of terms is reused.

13. Ontologies and data sets need to be updated automatically in order to provide the latest version and to document its development. The latter represents metadata.

14. Ontology Repositories and Data Catalogs should provide RDF links. Using well-known vocabularies, for example Dublin Core and FOAF, to describe metadata represent also a way of providing RDF links.

# 7 Multiple casestudy

## 7.1 BioPortal

BioPortal uses PURLs to identify ontologies. PURLs are Persistent Uniform Resource Locators and localize indirectly the resource since the underlying Web address can change through time. Therefore, continuity of references is provided[36]. BioPortal uses the *purl.bioontology.org* server to generate URIs if the ontologies are not provided with a URI. Those PURLs are dereferenced by using the HTTP mechanism, for instance http://purl.obolibrary.org/obo/DOID_299.

The ontologies are formalized using different representation languages. Beside OWL, RDF and RDFs ontologies, there are also ontologies formalized in the Open Biological and Biomedical Ontologies (OBO) format and the Rich Release Format (RRF).

Accessing knowledge it occurs via Web services, which are RESTful services [Fiel02]. After entering keywords, which include advanced options in order to find appropriate ontologies, BioPortal is crawled via RESTful services, which include getting entities of a specific concept and its details. The RESTful architecture consists of four words GET, POST, PUT and DELETE. The results are ranked according to the best and at most, 100 matches. Browsing ontologies is only possible within BioPortal. BioPortal does not provide a query endpoint. Two systems store ontologies: the Mayo Clinic's LexGrid system[37] stores OBO ontologies and Protégé[38] [NoRM] for OWL and RDF ontologies. Protégé is a tool for ontology development and knowledge-acquisition and provides further plugins as providing visualization and support queries.

In order to visualize ontologies, BioPortal uses RESTful services and the Protégé plugin for OWL and RDF ontologies.

Quality assessment is provided by the user. Reviews can be made by a 1-5 star rating, so that the ontology is rated according its domain coverage, correctness, quality of content and usability inter alia. Besides, a report of experience can be made. However, those quality assessments methods are subjective.

[36] http://purl.bioontology.org/docs/index.html
[37] https://wiki.nci.nih.gov/display/LexEVS/LexGrid+Background+Information
[38] http://protege.stanford.edu/

Mappings on BioPortal are between concepts and provided by different ways. Mappings can be added by registered users directly on BioPortal, they can originate from the ontology content provider or are generated automatically by algorithms, for example LOOM. Mappings are created by users and provided with metadata, which include also their provenance. Besides, users add notes. Mappings in BioPortal are of the following three types OWL, RDFS and SKOS[39]: `owl:sameAs`, `rdfs:seeAlso` and `skos:relatedMatch`, `skos:closeMatch`, `skos:exactMatch`, `skos:broadMatch`and`skos:narrowMatch`.

BioPortal provides metadata about the provenance of the ontology. Therefore, BioPortal uses the Resource Index API, which is available through a Restful web service. After the resource element is fetched, the Annotator web service annotates the resource with terms in the ontology. While searching for a specific ontology or concept, the user can request the resource index, which is linked to the original resource. BioPortal is an open ontology repository [BaSc09], but some ontologies are marked so that they are either private, or licensed. First of all, the user has to be registered in order to get access to those kinds of ontologies. Then, if the ontology is private, the user has to contact the submitter. If the ontology is licensed, the user is asked to enter the licensing information. Therefore, the access is slightly limited. Ontology metrics, as part of metadata about ontologies, are provided. Those consist of statistical metrics and quality-control and quality-assurance metrics. Reviews, details, versions of an ontology and the projects they are used represent metadata as well as mappings and added notes in order to discuss the ontology classes. The metadata is represented using the OMV vocabulary.BioPortal ontologies are registered on thedatahub.org, which is a data catalog [NSWD09], [WNSA11], [MuNC11].

[39] [MiBe09]

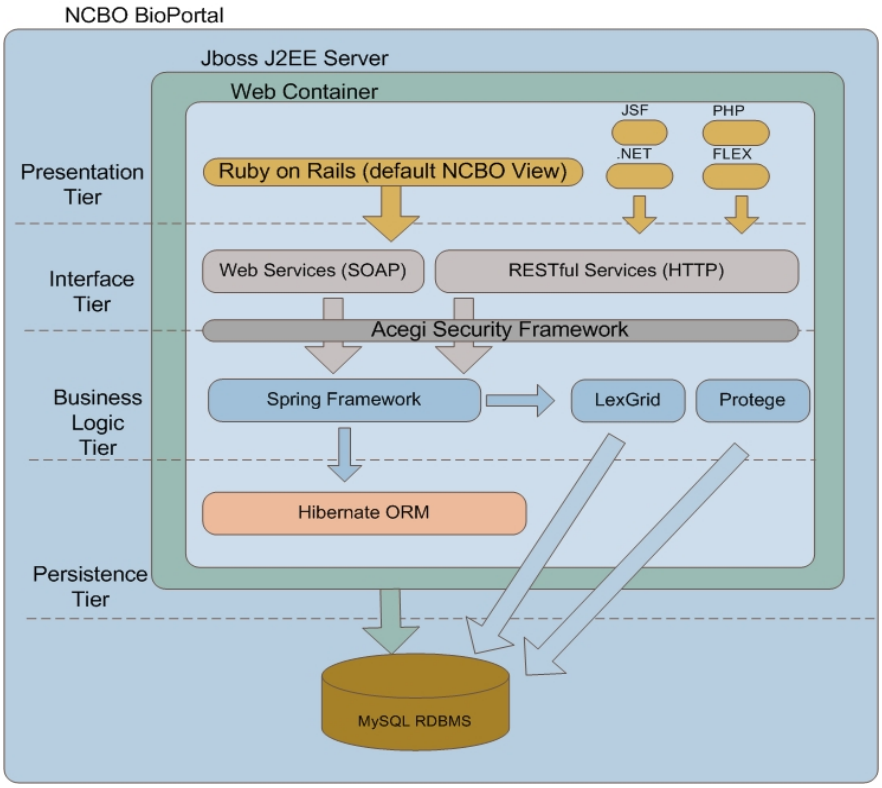



**Illustration 3[MuNC11]**

## 7.2 Cupboard

On Cupboard ontology spaces are created in order to submit ontologies. Ontologies are provided with URIs, which are dereferenced. Therefore, the user specifies the location on the filesystem where the ontology is stored directly to a URL. Then Cupboard suggests a namespace and adapts it in order to get a URI, which can be dereferenced. Cupboard provides OWL and additionally, alignments of the formats OWL, C-OWL and SWRL [AqEL09].

A core component of Cupboard is the Watson Client API, which uses SOAP services for searching, browsing, receiving metadata and querying. Watson indexes ontologies in Cupboard. Therefore, the user can find and browse ontology spaces on Cupboard and the repository in general. The search occurs via keyword and can be refined. Additionally, the Cupboard API enables the user to restrict searching to a certain ontology space. For querying, the repository provides SPARQL queries. Within the own ontology space visualization is provided by displaying an ontology and the number its triples, alignments, reviews and the star rating bar.

Quality assessment is offered, provided the user can assess via star ratings according to reusability, correctness, complexity, domain coverage and modeling of the ontology and explanations can be made on the ratings. Besides, a topic specific open rating system (TS-ORS) is implemented, which enables users to add trust and meta-trust statements about the reviews for the property, for the ontology or for everything. With a RESTful service of TS-ORS, an overall rating for a certain ontology can be found. The TS-ORS system ranks ontologies according to these reviews [LeAE10].

In order to map ontologies and concepts, Cupboard uses the Alignment Server. Therefore, adding and uploading alignments is possible. In order to add alignments, the user selects the ontologies, which should be mapped for this method. With the Alignment Format, alignments can be uploaded.

Metadata about the provenance can be submitted, the creator of the ontology and the location of the resource can be entered. Cupboard integrates the Oyster system to enter, store and register metadata and the ontology metadata vocabulary OMV[40] to manage it. Oyster is a peer-to-peer system, which is integrated in Cupboard to manage OMV. It provides the reuse of ontologies by representing a solution concerning the management of metadata and standardizes the process of adding information. Oyster extracts automatically information from an ontology because it provides RDFS, DAML+OIL and OWL and leaves place to add the missing parts. Besides this, Oyster supports the retrieval of ontologies by formulating queries. The user can search ontologies not only thorough keywords, but also by their means. Oyster creates a query using the terms, which describe ontologies. Since Oyster is a peer-to-peer system, the ontologies uploaded in Cupboard are accessible for the entire Oyster network, queries are routed through it [AqLe09], [LeAE10].

[40] http://omv2.sourceforge.net/

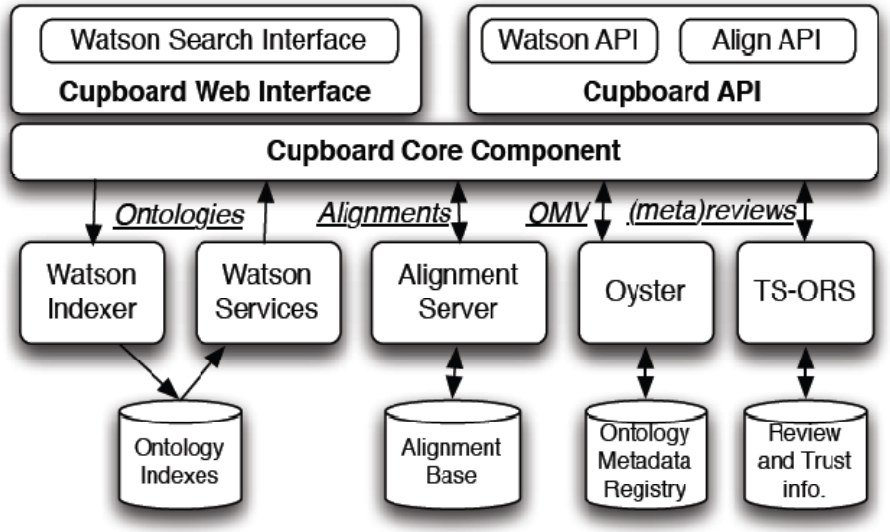



**Illustration 4[AqLe09]**

### 7.3 Watson

According to the illustration of the Watson architecture, the Watson search engine provides three main tasks, consisting of collecting, analyzing and querying the Web. Since Watson is a search engine for semantic documents, it does not provide URIs, but retrieves semantic documents, which are provided with a dereferenced URI.

Watson crawls, unlike the Watson API implemented in Cupboard, the entire Web in order to locate sources of semantic documents and collect ontologies and semantic documents. Because of this, sources such as the Protégé ontology library and Swoogle[41], a Semantic Web search engine, are crawled. A specialized crawler extracts sources by sending queries. A second crawler discovers new repositories and retrieves documents written in ontology languages. A third crawler retrieves already collected documents. Watson follows also mappings, for example owl:import and rdfs:seeAlso in order to retrieve information from other sources. The crawler set relies on Heritrix[42], an Internet Archive's open-source, extensible, web-scale, archival-quality web crawler project. Using Jena[43], the retrieved documents are eliminated, if those cannot be analyzed [ASDB07], [ABGS11].

Watson extracts and provides metadata about the entities contained in a document, their relations and the languages they are represented. Therefore, to validate the crawled information and then to analyze it, it is important to retrieve

[41] http://swoogle.umbc.edu/
[42] https://webarchive.jira.com/wiki/display/Heritrix/Heritrix
[43] http://jena.apache.org/

metadata since the case can arise that the same documents are provided with different URLs, or that different ontologies are provided with the same URI. After starting a search, the list of results provides already metadata [ASDB07]. At first, the URI of an ontology is shown, followed by the URIs containing the different entities. Then again, the URI of an entity contains ontologies, which describe the entity and explain which relations they have among each other, for example that one ontology is a subclass of another one. Navigating through the URIs, the metadata of a document can be retrieved. The metadata of a document is presented in tabulation. The list contains the size of the document and the number of statements, its representation languages, labels, comments, employed description language, the number classes, properties and individuals. Besides this, information about the author and a URI of the documents location is provided. Another type of metadata are reviews, therefore their number is noted in the list as well. Based on these elements of metadata, the value of an ontology is determined, so that an ontology is either semantically rich or simple structured. *Further* metadata are the semantic relations between ontologies, for example owl:import. In order to avoid providing the same metadata twice, the syntax of the documents are compared by a crawler and then compared again considering their semantic serialization and their representation languages. In the end, the combined results are indexed.

Reviewing is provided using Revyu, a generic reviewing and rating site. Revyu is a Linked Data-driven Web application, which consumes Linked Data by exploiting the interlinking with DBpedia[44] [Haus09]. DBpedia extracts structured data from Wikipedia and provides this data on the Web.

Watson provides keyword search, SPARQL queries and ontology exploration. Similar to common search engines, the user enters keywords in the blank and can specify the search options. The user can decide whether the matcher has to be exact so that the results match strictly the entered words, or whether a word matcher, which offers more results, is enough. Further, the user can choose the entities to be searched. These values are classes, properties and individuals. They can be combined in different ways or by selecting all three entities, so that a whole ontology is listed as a result. And the last one to decide is the scope, therefore what part of an entity should match the entered keyword.

[44] http://dbpedia.org/About

Then a list of results, with metadata, a list of entities and the location for each document, is shown. Besides this, the URI is also displayed. Through the URI it is possible to navigate and browse further semantic documents or rather ontologies. Selecting an entity, other ontologies are provided, which are identified by URIs, so that further browsing leads to the full view of an ontology. Since the Semantic Web is intended to enable humans as well as machines to browse the Web, so Watson has the same intention. Therefore, Watson provides also SPARQL queries, which are provided after selecting an ontology. An SPARQL endpoint is implemented by Jena and enables to perform a query, which is provided so far by one ontology [ABGS11], [ASDB07].

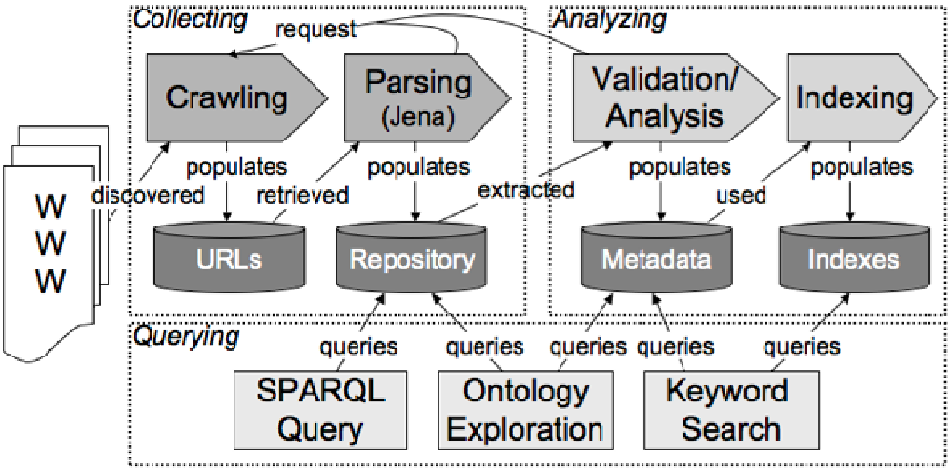



**Illustration 5[ABGS11]**

### 7.4 The Data Hub

The Data Hub is run by the open-source software CKAN, which enables searching, publishing, visualizing and versioning data sets and provides metadata. The contained data sets are identified by URLs and different representation formats are used. Beside RDF and SPARQL, also PDF and CSV are provided. The Data Hub stores metadata and links to the provenance data sets, but files, containing data sets can be uploaded or simply registered on the Data Hub.

The Data hub provides a full-text search, in which metadata attributes including tags and groups are searched. But there is also the possibility to set restrictions according to tags, groups and resources formats. Those metadata attributes

also facilitate browsing the data catalog, since users can follow tags or groups to retrieve similar data sets. Searching and querying are enabled through the RESTful JSON API.

Visualizing data sets is provided by the CKAN previewing tool, enabling different views, such as table views or graphing data, images or even a preview of the Web page of a resource. Additionally, data sets can be mapped if they contain details about their geographical location. To create preview views, the RESTful API is used [Smit12].

The quality assessment on the Data Hub occurs according the Linked Open Data star scheme suggested by Tim Berners-Lee [Bern11]. Therefore, the Multipurpose Internet Mail Extensions Type (MIME-Type), which identifies file formats on the Internet, of the data sets resources is reported. Based on the reported MIME-Type of each data set, the five star rating is calculated. The highest score are five stars; since these mean that the published data set provides RDF links and hence is Linked Data.

On the Data Hub, mappings are provided via tags.

Provenance metadata is provided by a link to the resource, but this is not provided for each data set. The Data Hub provides metadata for every data set. Metadata consist of a description about the data set and further information, such as the author and the maintainer of the data set, the version of the data sets and its state. The further information can vary. Since the data sets is editable, a revision history is provided with stating a certain time. Whether a data set provides a license is stated, so that users know if the data set is available. Besides, groups, which the data set belongs to, are also displayed [Smit12], [Smit].

### 7.5 OntoSelect

OntoSelect provides ontologies with Unique Resource Locators (URL), which address the location of ontologies. Ontologies are formalized using the representation languages RDF/S, OWL and the DARPA Agent Markup Language (DAML)[45]. DAML is a markup language based on RDF, which was further developed, so that much of it is incorporated into OWL.

[45] http://www.daml.org/

OntoSelect provides two ways of searching since it is also a search engine: a keyword-based search, a title-specific search and a topic-specific search. The latter enables the user to specify the search by either entering an URL or the topic itself, which is linked to Wikipedia, so that a matching web page appears. Using URLs, the exact ontology with its metadata is displayed. Collecting ontologies occurs on OntoSelect automatically by using the Google API[46]. The Google API crawls the Web and indexes each class and object property with reference to the ontology it is contained. Each class or property label is indexed as well as each ontology and the human language of the label. The ontologies are stored in a database according to the indexes. Therefore, OntoSelect can be browsed by ontology name, format, human language, class or property label and included ontologies. Further, searching and browsing occurs according to the number of classes and properties, the representation language and the connectedness.

OntoSelect provides ranking by coverage, structure and connectedness of ontologies. Therefore OntoSelect extracts all textual data and analyses this using the OWL API. After extracting all nouns, those are used in order to calculate the three factors. The coverage score indicates how many of the nouns are covered by the classes and properties in the ontology. The structure criterion is calculated by the number of properties of an ontology divided by the number of classes of the same ontology in order to determine the how detailed the knowledge structure of the ontology is. The third score measures if the ontology is connected to other ontologies and how well established those are. Since these three scores differ among themselves, a combined score is calculated, so that ranking occurs according to it. Therefore the most appropriate ontology can be found. A list, containing the 20 best matching ontologies is displayed [Buit04], [BuEi], [BuEi07], [BuED04].

### 7.6 Linked Open Vocabularies (LOV)

The Linked Open Vocabularies uses URIs to identify the contained vocabularies. Those must be dereferenced since LOV sets following requirements according to this: a vocabulary has to be retrievable by content negotiation from its namespace URI. Besides, a vocabulary has to be formalized using the representation languages RDFS or OWL. Therefore, the

[46] http://code.google.com/

vocabularies, which LOV contains, are machine-readable, so that machines are able to reason.

LOV provides searching using keywords. The search options can be refined according to domain and type, so that searching for properties and classes is possible. Querying is provided by a SPARQL endpoint, in which all vocabularies are aggregated. The search results are ranked according to metrics: element labels relevancy to the query string, element labels matched importance, number of element occurrences in the LOV data set, number of vocabulary in the LOV data set that refer to the element and number of element occurrences in the LOD. From these metrics a score is calculated, so that the results are ranked according to the score.

Since LOV aggregates all vocabularies in an endpoint, hence data is extracted in order to generate statistics. Those statistics are about vocabulary elements:

- The LOV distribution metric is about the number of vocabularies in LOV that refers to a particular element.
- LOV popularity metric is about the number of other vocabulary elements that refers to a particular one.
- LOD popularity metric is about the number of vocabulary element occurrence in the LOD.

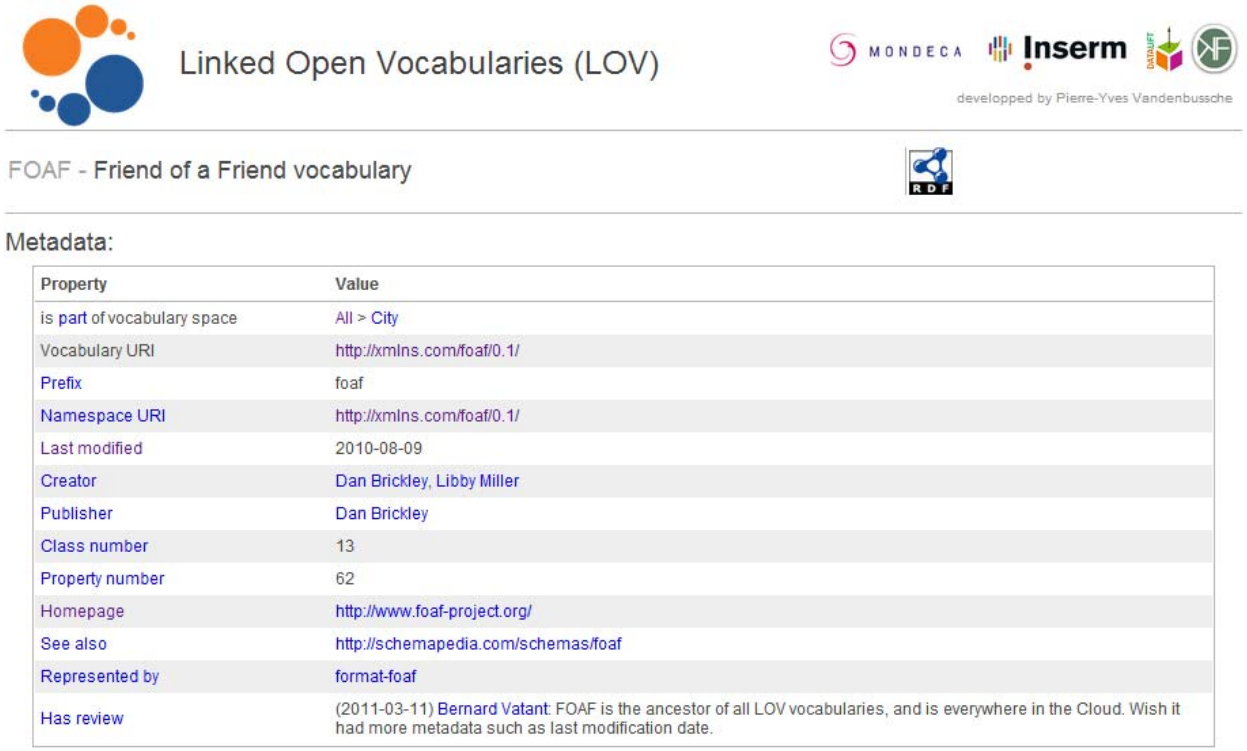
The provenance of a vocabulary is added as metadata, since its URI is stored in LOV. Further metadata is described, using different vocabulary metadata. LOV uses the Vocabulary of Interlinked Data sets (VoID)[47], a standard vocabulary for describing metadata about RDF data sets, Dublin Core Metadata Initiative (DCMI) Metadata Terms[48], which defines general metadata attributes such as title, subject and author and the Bibliographic Ontology (BIBO), which provides concepts and properties to describe citations and references [HeBi11]. Besides, LOV uses the Vocabulary of a Friend (VOAF)[49]. This vocabulary specification provides properties, classes and vocabularies allowing the description of vocabularies, which are used in the Linked Data Cloud. Therefore, vocabularies can be linked, so that they depend on each other and specified by extending and annotating them since VOAF provides describing metadata as properties.

[47] http://vocab.deri.ie/void/
[48] http://dublincore.org/documents/2012/06/14/dcmi-terms/?v=terms#
[49] http://lov.okfn.org/vocab/voaf/v2.0/index.html

Since LOV provides freely available content, the vocabularies are licensed under Creative Commons CC BY 3.0[50].

Since the last version of a vocabulary is checked daily, those are uploaded and imported in an endpoint using the LOV aggregator. The different versions of a vocabulary are visualized using a time line.

The vocabularies contained in LOV are interlinked among themselves. Those links are visualized. LOV itself reuses vocabularies such as Dublin Core and is mapped to the Linked Data Cloud.


Linked Open Vocabularies (LOV)

MONDECA Inserm

developped by Pierre-Yves Vandenbussche

FOAF - Friend of a Friend vocabulary

Metadata:

| Property | Value |
|---|---|
| is part of vocabulary space | All > City |
| Vocabulary URI | http://xmlns.com/foaf/0.1/ |
| Prefix | foaf |
| Namespace URI | http://xmlns.com/foaf/0.1/ |
| Last modified | 2010-08-09 |
| Creator | Dan Brickley, Libby Miller |
| Publisher | Dan Brickley |
| Class number | 13 |
| Property number | 62 |
| Homepage | http://www.foaf-project.org/ |
| See also | http://schemapedia.com/schemas/foaf |
| Represented by | format-foaf |
| Has review | (2011-03-11) Bernard Vatant: FOAF is the ancestor of all LOV vocabularies, and is everywhere in the Cloud. Wish it had more metadata such as last modification date. |

**Illustration 6 Metadata of the FOAF vocabulary provided in LOV http://lov.okfn.org/dataset/lov/details/vocabulary_foaf.html**

[50] http://creativecommons.org/licenses/by/3.0/

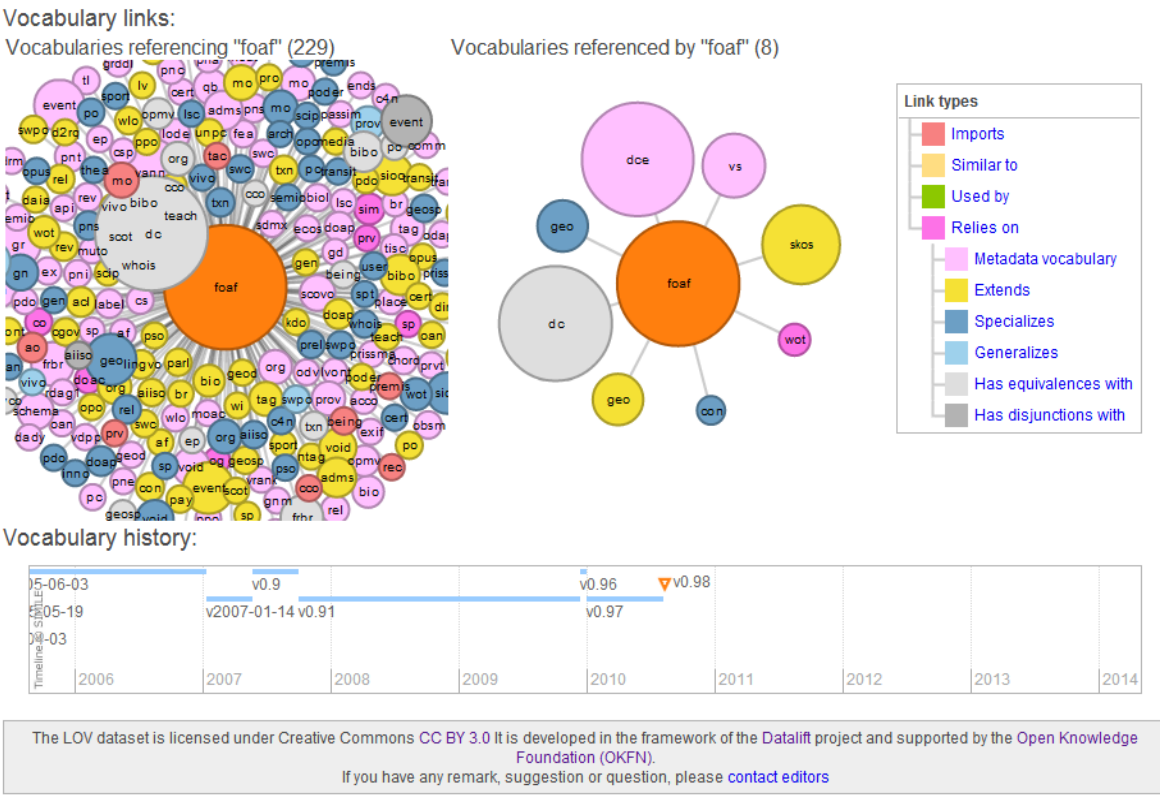



**Illustration 7Visualization of vocabulary links of FOAF within the LOD cloud and the vocabulary history provided by http://lov.okfn.org/dataset/lov/details/vocabulary_foaf.html**

### 7.7 DCAT

DCAT enables to express data catalogs as Linked Data. It is a RDF vocabulary and hence machine-readable, so that automated processing is possible. DCAT itself reuses other vocabularies, mainly the set of classes and properties defined by Dublin Core. Besides, FOAF terms are reused. With that, RDF links are already provided. Simultanously, a rich set of metadata is provided. DCAT provides four classes [MaEA12]:

- dcat:Data set: This class represents a data set using Dublin Core to describe metadata, such as title, location and time. Also which license is provided is described. The DCAT set describes the quality of the data sets.
- dcat:Catalog: represents a data catalog, which contains data sets.
- dcat:CatalogRecord: This class describes a data set entry, for example an update of the data set itself.
- dcat:Distribution: is a part of a data set, which is accessible by a web service or can be downloaded.

In order to make a data catalog Linked Data using DCAT, the class dcat:Catalog must be described by an author, who is provided with an URI through the FOAF vocabulary. SKOS describes conceptual hierarchies, therefore URIs are generated, which identify the data sets according the described domain using SKOS [MaEA12]. Consequently, using DCAT, data sets within data catalogs are provided with URIs.
DCAT can also be used, after a data catalog is already created and is able to improve it. For example, data catalogs, which may provide non-proprietary format [Bern11], such as CSV for instance but not RDF standards.

### 7.8 ONTOSEARCH2

ONTOSEARCH2 provides copies of ontologies represented in RDF, OWL DL its sublanguage OWL DL-Lite. DL-Lite can express features in UML class diagrams. Since OWL DL provides class and property constructions, the set of class and property axioms is called TBox and the set of individual axioms ABox. Out of the ABox and TBox the Relational Database Management System (RDBMS) is built. Therefore, the user is enabled to query either the ABox storage or the TBox. ONTOSEARCH2 provides a SPARQL endpoint in order to query the repository. Querying the ABox, the SPARQL queries must be parsed and converted to the DL-Lite conjunctive query format. This is converted into the Structured Query Language (SQL) again, since a RDBMS provides queries in SQL. The RDBMS returns a list of results using a results formatter, which returns HTML or XML, since ONTOSEARCH2 is implemented as a set of Java Servlets and JSP pages. TBox queries proceed similar to ABox queries. Reasoning is enabled by the OWL DL reasoned PELLET[51], which is a reasoner for Java. Retrieving ontologies is provided by a keyword-search within the repository. Browsing is not provided. Ontologies are submitted by users, who register either a URI or a URL of an ontology. Metadata and mappings of ontologies are not provided as well as metrics, rankings and reviews in order to assess the quality[PaTS].

[51] http://clarkparsia.com/pellet

| Ontology Repository / Linked data Requirements | Bioportal | Cupboard | Watson | The Data Hub | OntoSelect | LOV | DCAT | OntoSearch2 |
|---|---|---|---|---|---|---|---|---|
| Providing URIs | ✗ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ |
| Dereferenced http URIs | ✗ | ✓ | ✓ | ✗ | ✗ | ✓ | | |
| Machine-readable, standardized representation language | ✗ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ |
| Searching | ✗ | ✓ | ✓ | ✗ | ✓ | ✓ | ✓ | ✓ |
| Browsing | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✗ |
| Queryingendpoints | ✗ | ✓ | ✓ | ✗ | ✗ | ✓ | ✓ | ✓ |
| Visualization | ✓ | ✓ | ✗ | ✓ | ✗ | ✓ | ✗ | ✗ |
| Quality assessment | ✓ | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ | ✗ |
| Ranking data according quality assessment | ✗ | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ | ✗ |
| Mappings | ✓ | ✓ | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ |
| Provenancemetadata | ✓ | ✓ | ✗ | ✓ | ✗ | ✗ | ✓ | ✓ |
| Licensing metadata | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | | ✗ |
| Metadata | ✓ | ✓ | ✓ | ✓ | ✗ | ✓ | ✓ | ✗ |
| Automatic update ofversions | ✗ | ✗ | ✗ | ✗ | ✓ | ✗ | ✓ | ✗ |
| Linking | ✗ | ✗ | ✗ | ✗ | ✓ | ✓ | ✓ | ✗ |

### 7.9 Discussion

BioPortal provides ontologies with HTTP URIs, which are dereferenced, if those are represented using standardized machine-readable languages. But there are also OBO and RRF ontologies, so that BioPortal provides those ontologies with PURLs, which are also dereferenced. Therefore, one important requirement is already not fulfilled, namely to provide data expressed in formal specifications. Searching, browsing and querying are enabled via RESTful services, so that another Linked Data requirement is not met. Using the HTTP protocol to make calls between machines, humans can look up those names. The REST Web is a subset of the WWW and by its use, the full exploitation of the Semantic Web is enabled, since there are similarities between the REST API and Linked Data [PaRM11]. Additionally, BioPortal does not provide a set of crawlers, so that searching is currently limited within BioPortal. The implementation of Protégé tools in order to visualize the contained ontologies is efficient, since they enable to extract parts of the view and are designed for human consumption. Hence the Protégé plug-ins support to browse the ontologies and semi-automated merging and mapping [NoRM]. BioPortal provides only subjective quality assessment and ranks ontologies according the most matching keywords, further heuristics according the content or context would improve the quality of the contained ontologies. Ranking according matching keywords is not useful enough since users aim to retrieve ontologies of high quality. SKOS, RDF and OWL are expressive and their use is sufficient because reasoning is enabled. The Resource Index API provides provenance data; therefore the quality of the ontology can be assessed because the origin is known. Availability of ontologies can be restricted by their submitters, which limits other users, since licensing metadata is not provided and represents therefore a disadvantage. Ontologies on BioPortal need to be accessible in order to reuse them. Metadata is described with OMV, which simplifies retrieving and reusing ontologies. An automated update of ontologies is not provided, new versions must be uploaded. BioPortal is registered as a group on the Data Hub but not linked to it using RDF links.

Cupboard provides HTTP URIs, which are dereferenced via content negotiation and ontologies, which are machine-readable. Searching within the repository occurs through the REST Watson API, which does not crawl the entire Web, but

uses its search mechanism to retrieve ontologies within Cupboard and indexes them, which is a disadvantage, since retrieving appropriate ontologies is restricted as well as on BioPortal. Querying Cupboard is also provided by the Watson API using SPARQL queries, which can also be restricted to an ontology space, but providing only a SPARQL endpoint in order to query the entire Web is more appropriate than combining REST with SPARQL, since SPARQL stops REST to process appropriate [PaRM11]. Quality assessment is enabled with the TS-ORS and ranking occurs according to the TS-ORS. This heads towards Linked Data, however, the ratings are still subjective even if they are trust-rated. The alignment server facilitates to map ontologies and to store those mappings but users still have to align their ontologies by hand. Consequently, those alignments are only retrievable within Cupboard. Provenance metadata is submitted by the user. This implies that provenance metadata is not surely provided. Oyster is integrated to enter and store metadata, described with OMV and manages it. Besides, RDF links are not provided and common vocabularies to describe metadata are not used. Cupboard is a separate repository, although the Watson API is implemented.

Watson is a Semantic Web search engine and discovers ontologies, which are provided with a dereferenced HTTP URI and use standardized representation languages. Retrieving ontologies is enabled by a set of crawlers and SPAQRL queries. The crawlers retrieve sources of Semantic Web documents and follow also OWL and RDFS mappings. If ontologies cannot be parsed, they are eliminated. Besides, quality assessment occurs with Revya, a Linked Data application. The Watson search engine is a sufficient step towards Linked Data.

The Data Hub does not provide data sets as recommended with URIs, but with URLs. This is already clearly visible when users publish their data sets, since they are requested to add an URL. Moreover, not only machine-readable representation languages are used to describe data, but also data as PDF is provided by The Data Hub. Therefore reasoning of each data set is not possible. Also retrieving them can represent a problem, since URLs do not describe what the data set is about exactly. Querying and searching is enabled by the RESTful API. A SPARQL endpoint is not provided in order to query the Web, retrieving data sets is limited here to the data catalog itself. Browsing is facilitated using tags and groups, therefore similar data sets can

be retrieved. Visualizing data sets occurs through tools. Provenance metadata can be submitted by users by adding the origin link. Metadata about data sets consist of descriptions and additional information regarding data sets. Vocabularies, such as the Dublin Core or VoID, describing metadata are not provided. Hence, reusing the data sets for applications or retrieving them on the Web remains difficult. Similar data sets can be mapped within the Data Hub using tags or adding groups. An automatic update of data sets does not exist on the data catalog, but is editable by users. The quality assessment is based on the Linked Open Data star scheme, which illustrates, if a data set is Linked Data or not. The most data sets are not Linked Data, so that the Data Hub is not a Linked Data catalog, but an Open Data catalog, since also licenses are provided.

OntoSelect stores ontologies, which are provided with an URL and defined using DAML, RDFS and OWL. Therefore, ontologies are machine-readable and reasoning is enabled but using URLs is insufficient, since the location of the origin ontology can change and hence provenance metadata cannot be provided. Using the Google API OntoSelectcrawls the Web and retrieves ontologies, which are analyzed and indexed using the OWL API. This means, that OntoSelect does not restrict itself retrieving ontologies by searching only within the repository. A query endpoint is not provided. Measuring coverage, the structure and connectedness in order to calculate a score, serves as an appropriate method to rank ontologies. Simultaneously, these three criteria function as a quality assessment, since OntoSelect updates its repository dynamically, and enable users to select the most appropriate ontology. Nevertheless, OntoSelect is not an ontology repository, which provides Linked Data. That is because it does not provide RDF links and RDF standards such as a SPARQL endpoint and URIs to identify the collected ontologies; metadata is limited to human language, formats, domains, labels, number of classes and properties. Vocabularies to describe metadata are not used and mappings are not provided.

LOV represents a management tool for vocabularies, which are provided with URIs and PURLs. Those are dereferenced HTTP URIs and HTTP PURLs respectively, for example http://www.w3.org/2004/02/skos/core, and described in RDFS and OWL. Since security is provided and therefore

licensed under creative commons, the content is freely available. This indicates LOV as an open data catalog. Searching and querying is based on RDF standards, therefore LOV provides a SPARQL endpoint. However, querying is only within the catalog possible. Vocabularies are ranked according to metrics, which is calculated . Provenance of the vocabulary is provided by its URI or rather PURL, which leads to the origin of the vocabulary. Metadata about vocabularies is described by reusing existing terms VoID, Dublin Core and BIBO, so that applications can better consume metadata. In addition, VOAF is used to describe metadata. Links between vocabularies are visualized but not every vocabulary in LOV is linked to another one. Consequently, LOV does not fulfill the Linked Data requirements, although it is linked to the LOD cloud.

DCAT is very popular as a Linked Data vocabulary, since it is a standard RDF scheme vocabulary. Therefore, its use enables data catalogs to provide their data sets in machine-readable form. This enables data catalogs to provide advanced querying and retrieving methods. Additionally, DCAT provides RDF links, since it includes and hence, reuses other well-known vocabularies, which are recommended to use in order to provide Linked Data, such as SKOS, the Dublin Core and FAOF [HeBi11]. An example which illustrates the successful application of DCAT within data catalogs is [CyMP10]. Here, it is described, how DCAT is used in already existing data catalogs. Therefore, the data sets were stored in a relational database and then mapped to the DCAT vocabulary. Afterwards, the data sets were published as Linked Data and in combination with the D2R Server[52] [HeBi11] a SPARQL endpoint was provided.

Querying via SPARQL is enabled, although ONTOSEARCH2 provides ontologies formalized in OWL DL additionally. Therefore, SPARQL queries are converted into DL-Lite conjunctive queries; those are converted into SQL queries. This is possible, since OWL DL provides some extensions, which are constructors. ONTOSEARCH2 enables reasoning through PELLET. However, reasoning is still restricted, since the infer engine processes for the DL-Lite ontology language. ONTOSEARCH2 searches and queries only within the repository and not all of the ontologies are

[52] http://d2rq.org/d2r-server

provided with dereferenced HTTP URIs. The repository and the ontologies respectively are not interlinked using RDF links; metadata except the provenance is not at all provided, whereas provenance is only provided, if the user submits the origin URI of an ontology. Additionally, quality assessment and hence ranking mechanisms according the quality are not implemented, as well as visualizing tools such as the Protégé plug-in. Since most of the Linked Data requirements are not fulfilled and the reuse of the stored ontologies is more difficult than facilitated, ONTOSEARCH2 is not an ontology repository, which provides Linked Data.

The multiple case study carried out on representative ontology repositories and data catalogs shows that there is a lack of applying Linked Data. This is apparent from Linked Data requirements, which are not fulfilled. The ontology repositories are separate platforms, which address different user requirements. Retrieving an appropriate ontology or data set seems only possible, if the user accesses every repository one after another, so that simultaneous access is not possible. In addition, repositories and data catalogs do not provide the uniform identification of ontologies and data sets using URIs.Besides, different representations languages are used. Only the search engine Watson enables access to ontologies of different sources.

# 8 Vocab.cc

Vocab.cc is an open source project and part of the Planet Data project[53], which aims to enable researcher to provide their data in new and useful ways. Therefore vocab.cc enables to retrieve easily RDF data within the context of Linked Data. To prove this, a case study on vocab.cc is carried out, in order to find out, if the Linked Data requirements on ontology repositories and data catalogs are fulfilled by vocab.cc.

## 8.1 The Billion Triple Challenge Data set

The data set, which vocab.cc provides, origins from the Billion Triple Challenge Data set 2011[54] (BTCD). BTCD 2011 was crawled from the Web using a random sample of URIs from the BTCD 2010. The contained statements are formalized using N-Quads[55]. N-Quads is a format, which extends N-Triples[56] with additional information, so that the statements consist of triples in shape of `<subject><predicate><object><context>`in comparison to N-Triples, where as a triple consists of `<subject><predicate><object>.` The fourth part `<context>` provides provenance metadata about a data set. BTCD identifies vocabularies and delivers metadata about the domains those cover and their relevance. Relevance of vocabularies means here, how often those are used [Hart11], [Mend11].

## 8.2 Case study on vocab.cc

Access to information is provided directly on the user interface with a field in order to lookup or search a URI. These are the two main tasks, which vocab.cc provides, namely to lookup metadata by specifying a URI or to search for data by formulating a query regarding the user's interests. Dereferenced HTTP URIs are provided using content negotiation.

To lookup a URI, users specify a URI or its namespace. If these namespaces have common prefixes, prefix.cc[57] resolves them automatically, since prefix.cc supports to lookup URI prefixes. If the URI identifies a property or a class in the

[53] http://planet-data.eu/
[54] http://km.aifb.kit.edu/projects/btc-2011/
[55] http://sw.deri.org/2008/07/n-quads/
[56] [CyHH08]
[57] http://prefix.cc/

BTCD, the displayed metadata contains whether the URI describes a property or a class, the complete URI, the number of overall appearance in the BTCD and the number of appearance in data sets. According to the two kinds of appearances, there are two ranking lists: an overall ranking and a data set ranking, so that both positions in the rankings are also displayed.

Searching a URI in order to retrieve an appropriate vocabulary, which can describe the user's data set, works by formulating a simple query according the users interest. Therefore matching labels are found, so that URIs of possible vocabularies are displayed. Additionally, the number of appearance in the BTCD and the type of the URI is stated. For the individual URIs, the lookup functionality is integrated. The orange arrow button leads to the particular vocabulary.

Since SPARQL is provided for querying, vocab.cc is able to reason using the SPARQL graph patterns. The quality of a data set is represented by its relevance. This means, if the vocabulary is often used, the user can assume that it is well adopted and useful [Mend11].

| **Linked Data Requirements** | **Vocab.cc** |
|---|---|
| Providing URIs | URIs provided |
| Dereferenced HTTP URIs | Dereferenced HTTP via content negotiation |
| Machine-readable, standardized representation language | RDF |
| Searching and browsing | Linked services: lookup and search |
| Querying endpoints | SPARQL endpoint |
| Visualization | - |
| Quality assessment | Usage of data, |
| Ranking data according quality assessment | According usage of data, information provided by BTCD |
| Mappings | - |
| Provenance metadata | N-Quads: fourth node |
| Licensing metadata | Services: LGPLv3, content: CCBYSA |
| Metadata | - |
| Automatic update of versions | - |
| Linking | Prefix.cc |

### 8.3 Discussion

The use of dereferenced HTTP URIs is one of the Linked Data principles. Vocab.cc executes this principle via content negotiation, so that humans and machines are enabled to read a representation.

The URI lookup and the URI search can also be accessed as Linked Services which are a combination of RESTful services Linked Data. However, this reveals difficulties since there are differences in both architectures. This becomes apparent in resource identification. REST enables retrieving data and provides links in order to navigate. Standard representation languages identify resources to encapsulate data and the links build a web.Also the use of SPARQL is an impediment in combination with SPARQL [PaRM11].

Vocab.cc makes use of BTCD, which provides statements represented in N-Quads. N-Quads enables to provide provenance metadata, which is very useful since the origin of the data set is used to assess its quality and builds trust. Licensing metadata is provided as well. The data sets are licensed under CC BY-SA, so that the content can be legally used. Since N-Quads is an extension of RDF, the provided vocabularies are machine-readable, so that reasoning is enabled. Querying is provided by SPARQL across the BTCD.

Additionally, also the Linked Open Data star scheme can be carried out in order to point out whether vocab.cc is a Linked Data vocabulary catalog or not:

- Vocab.cc provides vocabularies on the Web, which are freely available under the CC BY-SA license; the source codes of the Linked Services are also licensed under LGPLv3.
- Vocab.cc provides RDF vocabularies, consequently these are machine-readable. Hence non-proprietary format is provided.
- The vocabularies are provided with RDF standards; except that vocab.cc provided RDF vocabularies, these are provided with URIs. The HTTP URIs are dereferenced via content negotiation, so that machines are able to reason and automatically either HTML documents are displayed or RDF data is delivered. Besides, vocab.cc queries are carried out via SPARQL.
- The last star requires providing RDF links: vocab.cc is interlinked with prefix.cc, which resolves automatically namespaces.

Therefore, also from the Linked Open Data star scheme can be concluded, that vocab.cc is Linked Data.

### 8.3.1 Recommendations

The case study and also the LOD star scheme evince, that vocab.cc meets the Linked Data requirements. However, recommendations can be made in order to improve vocab.cc. A richer set of metadata should be provided, for instance information about the size, the structure and the author. In addition, the multiple case study shows that the visualization of data sets and ontologies facilitates users, who are not experts, selecting appropriate data. Therefore, vocab.cc should implement visualizing tools, so that different groups of users are supported equally. Since the BTCD is crawled once a year[58], an automated update of data sets is not provided. However, this would improve the quality of data sets. Additionally, retrieving vocabularies should be provided across the entire Web and not restricted according the BTCD. This would raise the number of retrieved vocabularies. Therefore, quality assessment methods need to be provided directly on vocab.cc in order to assure quality furthermore.

## 9 Outlook

The multiple case study illustrates that the representative ontology repositories and data catalogs, which are selected, do not fulfill the Linked Data requirements. But the developers of particular platforms are aiming to provide their repositories and catalogs with Linked Data technology.

BioPortal [SAMN09], [LNSW10], for instance, is starting to translate the contained ontologies into RDF data and is going to provide these versions at http://sparql.bioontology.org, which is a beta version. Since ontologies on BioPortal are represented in different languages beside OWL and RDF, URIs are generated for those using the purl.bioontology.org server according to the following convention:

[58] according the Web page of the BTCD 2011

- OWL and RDF/S ontologies: URIs defined in the ontologies
- Protégé Frames & RRF Ontologies: http://purl.bioontology.org/ontology/{abbreviation}/{concept id}
- OBO ontologies: http://purl.obolibrary.org/obo/{idspace}_{localid}.

Simultaneously, the use of RDF enables querying via SPARQL. Therefore, a SPARQL endpoint is provided, so that ontologies are retrievable either as a whole, or as specific RDF graph patterns. Furthermore, REST services enable access and download ontologies, which are the latest versions. Mappings between terms can be submitted by users through REST APIs or the Web interface. An effort is made so that mappings can be generated automatically. SKOS-based relationships are used.

As the case study on vocab.cc demonstrates, the effort of Linked Data contributes to facilitate retrieving appropriate ontologies.

## 10 References

[ABGS11] Aquin, M. de et al.: Watson: supporting next generation semantic web applications, Villareal, Spain, 2011.

[AqEL09] Aquin, M. de; Euzenat, J.; Lewen, H.: Sharing and reusing aligned ontologies with Cupboard.

[AqLe09] Aquin, M. de; Lewen, H.: Cupboard – A Place to Expose your Ontologies to Applications and the Community. ESWC: ESWC. Springer, 2009; S. 913–918.

[ASDB07] Aquin, M. de et al.: WATSON: a gateway for the semantic web, 2007.

[BaSc09] Baclawski, K.; Schneider, T. Hrsg.: The Open Ontology Repository Initiative: Requirements and Research Challenges, 2009.

[BCCC06] Berners-lee, T. et al.: Tabulator: Exploring and analyzing linked data on the semantic web.In Proceedingsof the 3RD internationalSemantic Web user interaction workshop, 2006.

[BeHL01] Berners-lee, T.; Hendler, J.; Lassila, O.: The Semantic Web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities. In Scientific American, 2001, 284; S. 34–43.

[Bern11] Berners-lee, T.: Linked Data - Design Issues. http://www.w3.org/DesignIssues/LinkedData.html, 03.08.2012.

[BiCH07] Bizer, C.; Cyganiak, R.; Heath, T.: How to publish Linked Data on the Web. http://www4.wiwiss.fu-berlin.de/bizer/pub/linkeddatatutorial/, 08.08.2012.

[BiHB09] Bizer, C.; Heath, T.; Berners-lee, T.: Linked Data - The Story So Far. In International Journal on Semantic Web and Information Systems, 2009, 5; S. 1–22.

[BuED04] Buitelaar, P.; Eigner, T.; Declerck, T.: OntoSelect: A dynamic ontology library with support for ontology selection: In Proceedings of the Demo Session at the International Semantic Web Conference.

[BuEi] Buitelaar, P.; Eigner, T.: Ontology search with the OntoSelect Ontology Library. http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.177.6052.

[BuEi07] Buitelaar, P.; Eigner, T.: Evaluating Ontology Search. http://ceur-ws.org/Vol-329/paper02.pdf, 22.08.2012.

[Buit04] Buitelaar, P.: OntoSelect: Towards the Integration of an Ontology Library, Ontology Selection and Knowledge Markup: Proc. of the Workshop on Knowledge Markup and Semantic Annotation (Semannot2004) at the International Semantic Web Conference, 2004.

[CaSh06] Cardoso, J.; Sheth, A.P. Hrsg.: Semantic Web Services, Processes and Applications. Springer Science+Business Media, New York, New York, 2006.

[CyHH08] Cyganiak, R.; Harth, A.; Hogan, A.: N-Quads: Extending N-Triples with Context. http://sw.deri.org/2008/07/n-quads/, 27.09.2012.

[CyJe11] Cyganiak, R.; Jentzsch, A.: Linked Open Data cloud diagram. http://richard.cyganiak.de/2007/10/lod/imagemap.html, 18.08.2012.

[CyMP10] Cyganiak, R.; Maali, F.; Peristeras, V. Hrsg.: Self-Service Linked Government Data with dcat and Gridworks. ACM, New York, NY, 2010.

[DaRo11] Dadzie, A.-S.; Rowe, M.: Approaches to Visualising Linked Data: A Survey. In Challenges, 2011; S. 1–2.

[DeAn12] Dengel; Andreas Hrsg.: Semantische Technologien. Grundlagen - Konzepte - Anwendungen. SpektrumAkademischerVerlag, Heidelberg, 2012.

[DiFe01] Ding, Y.; Fensel, D.: Ontology Library Systems: The key to successful Ontology Reuse, 2001.

[Fiel02] Fielding, R. T.: Architectural Styles and the Design of Network-based Software Architectures. Dissertation, Irvine, California, 2002.

[Grub93] Gruber, T. R.: A translation approach to portable ontologies. In Knowledge Acquisition, 1993, 5; S. 199–220.

[HaPG09] Hartmann, J.; Palma, R.; Gómez-Pérez, A.: Handbook on Ontologies. Ontology Repositories.Springer Berlin Heidelberg, 2009.

[Hart11] Harth, A.: Billion Triple Challenge 2011 Dataset. http://km.aifb.kit.edu/projects/btc-2011/, 27.09.2012.

[Haus09] Hausenblas, M.: Linked Data Application - The Genesis and the Challenges of Using Linked Data on the Web. http://linkeddata.deri.ie/sites/linkeddata.deri.ie/files/lod-app-tr-2009-07-26_0.pdf, 04.08.2012.

[HeAn05] Heery, R.; Anderson, S.: Digital Repositories Review. http://www.jisc.ac.uk/uploaded_documents/digital-repositories-review-2005.pdf, 20.08.2012.

[HeBi11] Heath, T.; Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool, 2011.

[KDFO] Klein, M. et al.: Ontology Management: Storing, Aligning, and Maintaining Ontologies. In (Davies, J.; Fensel, D.; van Harmelen, F. Hrsg.): Towards the Semantic Web: Ontology-driven Knowledge ManagementTowards the Semantic Web. Wiley, Chichester, UK, 2003; S. 47–69.

[LeAE10] Lewen, H.; Aquin, M. de; Elahi, S.: Cupboard - Supporting ontology reuse by combining a Semantic Web gateway, Ontology Registry and Open Ratings Systems. Improved and final version. http://www.neon-project.org/nw/images/f/f5/NeOn_2010_D146.pdf, 27.09.2012.

[LHLS04] Lara, R. et al.: An evaluation of Semantic Web portals, 2004.

[LiMe11] Linckels, S.; Meinel, C.: E-Librarian Service. User-Friendly Semantic Search in Digital Libraries.Springer Berlin Heidelberg, 2011.

[Link11] Linked Data community Linked Data community: W3C Mailing-List archives. Mailingliste der Linked Data und Semantic Web community; Betreff: data schema / vocabulary / ontology / repositories, 2011.

[LNSW10] LePendu, P.; Noy, N. F.; Shah, N. H.; Whetzel, P. L.; Musen, M. A.: Exposing BioPortal Ontologies as Linked Data. http://www.stanford.edu/~plependu/ismb10.pdf, 22.08.2012.

[MaEA12] Maali, F.; Erickson, J.; Archer, P.: Data Catalog Vocabulary (DCAT). http://www.w3.org/TR/vocab-dcat/#introduction.

[Mcva04] McGuinness, D. L.; van Harmelen, F.: OWL Web Ontology Language. http://www.w3.org/TR/owl-features/.

[Mend11] Mendes, P. N.: Planet Data Deliverable D4.1. PlanetData data sets, vocabularies and provisioning tolls catalogue and access portal. http://www.planet-data.eu/sites/default/files/pr-material/deliverables/D4.1_Data_sets,_vocabularies_and_provisioning_tools_catalogue_and_access_portal.pdf, 27.09.2012.

[MiBe09] Miles, A.; Bechhofer, S.: SKOS Simple Knowledge Organization System Namespace. http://www.w3.org/2009/08/skos-reference/skos.html.

[MuNC11] Musen MA, N. N. S. N. W. P. C. C. S. M. S. B.; NCBO team: NCBO Developer Documentation - NCBO Wiki. http://www.bioontology.org/wiki/index.php/NCBO-OOR_Architecture, 27.09.2012.

[NoRM] Noy, N. F.; Rubin, D. L.; Musen, M. A.: Making Biomedical Ontologies and Ontology Libraries Work. http://bmir.stanford.edu/file_asset/index.php/1035/BMIR-2004-1131.pdf.

[NSWD09] Noy, N. F. et al.: BioPortal: ontologies and integrated data resources at the click of a mouse. In Nucleic Acids Research, 2009, 37; S. W170.

[PaRM11] Page, K. R.; Roure, D. C. de; Martinez, K.: REST and Linked Data: A match made for domain driven development? In (Alarcón, R.; Pautasso, C.; Wilde, E. Hrsg.): Proceedings of the Second International Workshop on RESTful Design, 2011; S. 22–25.

[PaTS] Pan, J. Z.; Thomas, E.; Sleeman, D.: ONTOSEARCH2: Searching and querying Web ontologies. In (Nunes, M. B.; Isaías, P.; Martínez, I. J. Hrsg.): Proceedings of the IADIS International Conference WWW/Internet 2006, 2006; S. 211–218.

[PrSe08] Prud'hommeaux, E.; Seaborne, A.: SPARQL Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/, 27.09.2012.

[SaCy08] Sauermann, L.; Cyganiak, R.: Cool URIs for the Semantic Web. http://www.w3.org/TR/cooluris/, 18.08.2012.

[SAMN09] Salvadores, M. et al.: BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. In Semantic Web Journal (IOS Press), 2009.

[Smit] Smith, S.: CKAN Features | ckan - The open source data portal software. http://ckan.org/features/, 27.09.2012.

[Smit12] Smith, S.: Opening up scientific data with CKAN and the DataHub | Open Knowledge Foundation Blog. http://blog.okfn.org/2012/06/19/ckan-science/, 27.09.2012.

[TaAS10] Tartir, S.; Arpinar, I. B.; Sheth, A. P.: Ontological evaluation and validation, 2010.

[TSVH10] Tuominen, J. et al. Hrsg.: A User Interface for Ontology Repositories. CEUR Workshop Proceedings, 2010.

[WNSA11] Whetzel, P. L. et al.: BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. In Nucleic Acids Research, 2011, 39; S. 541–545.

[Worl04] World Wide Web Consortium W3C: RDF Resource Description Framework. http://www.w3.org/RDF/.

[Zvie10] Zviedris, M. Hrsg.: Ontology Repository for User Interaction. CEUR Workshop Proceedings, 2010.